CS 726: Homework #1

Posted: 01/28/2020, due: 02/10/2020 in class

Please typeset or write your solutions neatly! If we cannot read it, we cannot grade it.

Q 1. Prove that all isolated minima are strict.
Hint: One way to prove this is by contrapositive. [7pts]

Solution. First, if $\mathbf{x}^*$ is not a local minimizer then trivially it cannot be an isolated local minimizer.

Suppose that $\mathbf{x}^*$ is a local minimizer, but it is not strict. Then, there exists a neighborhood $\mathcal{N}_{\mathbf{x}^*}$ such that for all $\mathbf{x} \in \{\mathcal{N}_{\mathbf{x}^*} \cap \mathrm{dom}(f)\} : f(\mathbf{x}) \geq f(\mathbf{x}^*)$. Further, because $\mathbf{x}^*$ is not strict, every neighborhood of $\mathbf{x}^*$ will contain a point $\bar{\mathbf{x}}$ such that $f(\bar{\mathbf{x}}) = f(\mathbf{x}^*)$. But, then, every neighborhood of $\mathbf{x}^*$ contains another local minimum $(\bar{\mathbf{x}})$ and, thus, $\mathbf{x}^*$ is not isolated. □

Q 2. A saddle point is a stationary point (a point $\mathbf{x}$ with $\nabla f(\mathbf{x}) = 0$) that is neither a local minimum nor a local maximum. List all stationary points of the following functions, where $\mathbf{x} = \left[\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right]$, and for each stationary point indicate whether it is a local minimum/maximum, a global minimum/maximum, or a saddle point.

(i) $f(\mathbf{x}) = \frac{{x_1}^4}{20} - 10{x_1}^2 + 10{x_2}^2$; [5pts]

(ii) $f(\mathbf{x}) = {x_1}^2 + {x_2}^2 + \frac{1}{\pi}\big(\sin(\pi \cdot {x_1}^2) + \cos(\pi \cdot {x_2}^2)\big)$; [5pts]

(iii) $f(\mathbf{x}) = \frac{1}{2}{x_1}^2 + x_1 \cos(x_2)$. [5pts]

Solution.
(i) Observe that:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} \frac{x_1}{5}({x_1}^2 - 100) \\ 20x_2 \end{bmatrix} \quad \text{and} \quad \nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{3}{5}{x_1}^2 - 20 & 0 \\ 0 & 20 \end{bmatrix}.$$

We have that $\nabla f(\mathbf{x}) = 0$ at points $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$, $\left[\begin{smallmatrix} -10 \\ 0 \end{smallmatrix}\right]$, and $\left[\begin{smallmatrix} 10 \\ 0 \end{smallmatrix}\right]$.

Point $\mathbf{x} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ is a saddle point, because $\nabla^2 f(\mathbf{x})$ has both positive and negative eigenvalues (-20 and 20). Points $\left[\begin{smallmatrix} -10 \\ 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 10 \\ 0 \end{smallmatrix}\right]$ are both local minima, because for either of the two points both eigenvalues of $\nabla^2 f(\mathbf{x})$ are positive (equal to 40 and 20). These minima are also global.
(ii) Observe the gradient of f :

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1(1+\cos(\pi {x_1}^2)) \\ 2x_2(1-\sin(\pi {x_2}^2)) \end{bmatrix}.$$

The stationary points are obtained when $x_1 = 0$ or $\cos(\pi {x_1}^2) = -1$ and $x_2 = 0$ or $\sin(\pi {x_2}^2) = 1$. Equivalently, the stationary points are: 1) $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$, 2) $\begin{bmatrix} 0 \\ \pm\sqrt{\frac{(2k+1)}{2}} \end{bmatrix}$, 3) $\left[\begin{smallmatrix} \pm\sqrt{2k+1} \\ 0 \end{smallmatrix}\right]$, and 4) $\begin{bmatrix} \pm\sqrt{2k+1} \\ \pm\sqrt{\frac{(2k+1)}{2}} \end{bmatrix}$, where $k \in \mathbb{Z}_+$.

Let us first argue that $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ is a global maximum (it is also possible to argue that it is unique). By Taylor's theorem, there exists $\gamma \in (0, 1)$ such that:

$$\begin{aligned} f(\bar{\mathbf{x}}) &= f(\mathbf{x}) + \langle \nabla f(\mathbf{x} + \gamma(\bar{\mathbf{x}} - \mathbf{x})), \bar{\mathbf{x}} - \mathbf{x} \rangle \\ &= f(\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]) + \langle \nabla f(\gamma \bar{\mathbf{x}}), \bar{\mathbf{x}} \rangle \\ &= f(\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]) + 2\gamma \bar{x}_1^2 \Big(1 + \cos(\pi \gamma^2 \bar{x}_1^2)\Big) + 2\gamma \bar{x}_2^2 \Big(1 - \sin(\pi \gamma^2 \bar{x}_2^2)\Big). \end{aligned}$$

As both sine and cosine functions are bounded between -1 and 1 and $\bar{x}_1^2 \geq 0$, $\bar{x}_2^2 \geq 0$, it follows that $f(\bar{\mathbf{x}}) \geq f(\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right])$, $\forall \bar{\mathbf{x}}$, and thus $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ is a global minimum.

All the remaining points are saddle points. This can be concluded by looking at the Taylor approximation in their neighborhoods. Suppose that $\bar{\mathbf{x}} = \mathbf{x} + \Delta\mathbf{x}$, for some small $|\Delta x_1|$, $|\Delta x_2|$. Use Taylor's Theorem again to show that for some $\gamma \in (0, 1)$:

$$\begin{aligned} f(\bar{\mathbf{x}}) - f(\mathbf{x}) = &\, 2x_1\Delta x_1 \left(1 + \cos\left(\pi(x_1 + \gamma\Delta x_1)^2\right)\right) + 2\gamma(\Delta x_1)^2 \left(1 + \cos\left(\pi(x_1 + \gamma\Delta x_1)^2\right)\right) \\ &+ 2x_2\Delta x_2 \left(1 - \sin\left(\pi(x_2 + \gamma\Delta x_2)^2\right)\right) + 2\gamma(\Delta x_2)^2 \left(1 - \sin\left(\pi(x_2 + \gamma\Delta x_2)^2\right)\right). \end{aligned} \tag{1}$$

Unless $x_1 = 0$, the first term in the first line dominates the second term, and so:

$$\begin{aligned} &2x_1\Delta x_1 \left(1 + \cos\left(\pi(x_1 + \gamma\Delta x_1)^2\right)\right) + 2\gamma(\Delta x_1)^2 \left(1 + \cos\left(\pi(x_1 + \gamma\Delta x_1)^2\right)\right) \\ &\qquad\qquad \approx 2x_1\Delta x_1 \left(1 + \cos\left(\pi(x_1 + \gamma\Delta x_1)^2\right)\right). \end{aligned}$$

As $1 + \cos\left(\pi(x_1 + \gamma\Delta x_1)^2\right) \geq 0$, the sign of the entire term depends on the sign of $x_1\Delta x_1$, which can be either positive or negative, depending on how we choose Δx_1. We can make the same argument about the terms depending on x_2 in Eq. (1). As we can have both $f(\bar{\mathbf{x}}) \geq f(\mathbf{x})$ and $f(\bar{\mathbf{x}}) \leq f(\mathbf{x})$ in the neighborhood of each stationary point $\mathbf{x}$, it follows that each of the stationary points from 2), 3), 4) are saddles.

(iii) Observe the gradient and the Hessian of f :

$$\nabla f(\mathbf{x}) = \begin{bmatrix} x_1 + \cos(x_2) \\ -x_1 \sin(x_2) \end{bmatrix} \quad \text{and} \quad \nabla^2 f(\mathbf{x}) = \begin{bmatrix} 1 & -\sin(x_2) \\ -\sin(x_2) & -x_1\cos(x_2) \end{bmatrix}.$$

There are two ways to make the gradient zero: 1) by choosing $x_1 = 0$ and $\cos(x_2) = 0$ and 2) by choosing $x_1 = -\cos(x_2)$ and $\sin(x_2) = 0$.

If $x_1 = 0$ and $\cos(x_2) = 0$, then $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ \frac{2k+1}{2}\pi \end{bmatrix}$ for $k \in \mathbb{Z}$. Depending on whether k is even or odd, the Hessian matrix equals:

$$\begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}.$$

Either way, the determinant of the Hessian is negative (equal to -1), and thus the two eigenvalues must have opposite signs (as the determinant of a matrix equals the product of its eigenvalues). This means that the Hessian has both negative and positive eigenvalues, and so each of these stationary points is a saddle.

If $x_1 = -\cos(x_2)$ and $\sin(x_2) = 0$, then either $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 2k\pi \end{bmatrix}$ or $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2(k+1)\pi \end{bmatrix}$ for $k \in \mathbb{Z}$. Either way, the Hessian is the identity matrix, and so each one of these stationary points is a local minimum. At each of these local minima, $f(\mathbf{x}) = \frac{1}{2}\cos^2(x_2) - \cos^2(x_2) = -\frac{1}{2}\cos^2(x_2) = -\frac{1}{2}$, and, thus, each of these local minima is also global. $\square$

Q 3. Let $\mathbf{A}$ be a real symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Prove that, $\forall \mathbf{x} \in \mathbb{R}^n$:

(i) $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq \lambda_1\|\mathbf{x}\|_2^2$; [5pts]

(ii) $\mathbf{x}^T\mathbf{A}\mathbf{x} \leq \lambda_n\|\mathbf{x}\|_2^2$. [5pts]

Solution. If $\mathbf{x}$ is a vector of all zeros, both statements hold trivially, so assume that it is not. By a simple rescaling of both sides in (i) and (ii) by $\|\mathbf{x}\|_2^2$, it suffices to prove the statements for $\|\mathbf{x}\|_2 = 1$.

Using the spectral theorem, there exist orthonormal vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ such that $\mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{u}_i\mathbf{u}_i^T$. Hence:

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^n \lambda_i(\mathbf{x}^T\mathbf{u}_i)^2.$$

It remains to use that:

$$\lambda_1 = \Big(\min_{1\leq i\leq n} \lambda_i\Big)\Big(\sum_{i=1}^n (\mathbf{x}^T\mathbf{u}_i)^2\Big) \leq \sum_{i=1}^n \lambda_i(\mathbf{x}^T\mathbf{u}_i)^2 \leq \Big(\max_{1\leq i\leq n} \lambda_i\Big)\Big(\sum_{i=1}^n (\mathbf{x}^T\mathbf{u}_i)^2\Big) = \lambda_n.$$

$\square$

Q 4. Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{-\infty, +\infty\}$ be an extended real valued *convex* function. Assume that f is such that for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $f(\mathbf{y}) \leq \underline{\lim}_{\alpha\downarrow 0} f(\alpha\mathbf{x} + (1-\alpha)\mathbf{y})$. Prove that:

(i) If there exists a point $\mathbf{x}$ such that $f(\mathbf{x}) = -\infty$, then f is not real-valued anywhere – it equals either $-\infty$ or $+\infty$ everywhere. [5pts]

(ii) If, $\forall \mathbf{x} \in \mathbb{R}^n$, $|f(\mathbf{x})| \leq M$, for some constant $M < \infty$, then f must be a constant function (i.e., taking the same value for all $\mathbf{x} \in \mathbb{R}^n$). [5pts]

Solution. (i) Given $\mathbf{x} \in \mathbb{R}^n$ such that $f(\mathbf{x}) = -\infty$, assume f.p.o.c. that there exists $\mathbf{y} \in \mathbb{R}^n$ such that $f(\mathbf{y})$ is real-valued. As f is convex, for any $\alpha \in (0,1)$, $f(\alpha\mathbf{x}+(1-\alpha)\mathbf{y}) \leq \alpha f(\mathbf{x})+(1-\alpha)f(\mathbf{y}) = -\infty$. Taking the limit $\alpha \downarrow 0$, we get that $f(\mathbf{y}) \leq \underline{\lim}_{\alpha\downarrow 0} f(\alpha\mathbf{x}+(1-\alpha)\mathbf{y}) = -\infty$, which contradicts the assumption that $f(\mathbf{y})$ was real-valued.
(ii) Suppose f.p.o.c. that f is not constant. Then there exist $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ such that $f(\mathbf{x})-f(\mathbf{y}) > 0$. Let $\beta = f(\mathbf{x})-f(\mathbf{y})$. For any $\alpha \in (0,1)$, there exists $\mathbf{z}$ ($\mathbf{z} = \frac{\mathbf{x}-(1-\alpha)\mathbf{y}}{\alpha} \in \mathbb{R}^n$) such that $\mathbf{x} = \alpha\mathbf{z} + (1-\alpha)\mathbf{y}$. By convexity:

$$f(\mathbf{x}) \leq \alpha f(\mathbf{z}) + (1-\alpha)f(\mathbf{y}).$$

Rearranging the last inequality:

$$f(\mathbf{z}) \geq f(\mathbf{y}) + \frac{f(\mathbf{x}) - f(\mathbf{y})}{\alpha} \geq -M + \frac{\beta}{\alpha}.$$

But for any $\alpha < \frac{\beta}{2M}$ we would then have $f(\mathbf{z}) > M$, which is a contradiction. □

Q 5 (Jensen's Inequality)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Prove that for any sequence of numbers $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k \in \mathbb{R}^n$ and any sequence of non-negative scalars $\alpha_1, \alpha_2, \ldots, \alpha_k \geq 0$ such that $\sum_{i=1}^k \alpha_i = 1$ we have:

$$f\Big(\sum_{i=1}^k \alpha_i \mathbf{x}_i\Big) \leq \sum_{i=1}^k \alpha_i f(\mathbf{x}_i). \qquad [8pts]$$

Solution. The proof is by induction on k. The base case $k = 2$ holds by convexity of f. Assume that the claim is true for $k > 2$. Write $\sum_{i=1}^{k+1} \alpha_i\mathbf{x}_i = (1-\alpha_{k+1})\sum_{i=1}^k \frac{\alpha_i}{1-\alpha_{k+1}}\mathbf{x}_i + \alpha_{k+1}\mathbf{x}_{k+1}$, and observe that $\sum_{i=1}^k \frac{\alpha_i}{1-\alpha_{k+1}} = 1$. We have:

$$\begin{aligned}
f\Big(\sum_{i=1}^{k+1} \alpha_i\mathbf{x}_i\Big) &= f\Big((1-\alpha_{k+1})\sum_{i=1}^k \frac{\alpha_i}{1-\alpha_{k+1}}\mathbf{x}_i + \alpha_{k+1}\mathbf{x}_{k+1}\Big)\\
&\leq (1-\alpha_{k+1})f\Big((1-\alpha_{k+1})\sum_{i=1}^k \frac{\alpha_i}{1-\alpha_{k+1}}\mathbf{x}_i\Big) + \alpha_{k+1}f(\mathbf{x}_{k+1})\\
&\leq (1-\alpha_{k+1})\sum_{i=1}^k \frac{\alpha_i}{1-\alpha_{k+1}}f(\mathbf{x}_i) + \alpha_{k+1}f(\mathbf{x}_{k+1})\\
&= \sum_{i=1}^{k+1} \alpha_i f(\mathbf{x}_i),
\end{aligned}$$

where the first inequality is by convexity of f and the second inequality is by the inductive hypothesis. □

Q 6. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex continuously differentiable function. Using the definition of convexity from the class and properties of directional derivatives, prove that it must be $\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle. \qquad [10pts]$$

Solution. If f is convex, then for all $\alpha \in (0,1)$:

$$f((1-\alpha)\mathbf{x} + \alpha\mathbf{y}) = f(\mathbf{x} + \alpha(\mathbf{y}-\mathbf{x})) \leq (1-\alpha)f(\mathbf{x}) + \alpha f(\mathbf{y}).$$

Rearranging:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \frac{f(\mathbf{x}+\alpha(\mathbf{y}-\mathbf{x})) - f(\mathbf{x})}{\alpha}.$$

It remains to take $\alpha \downarrow 0$. □

Q 7. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice continuously differentiable function. Prove that if f is convex, then it must be

$$\nabla^2 f(\mathbf{x}) \succeq 0, \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

that is, the Hessian of f is positive semidefinite. [10pts]

Solution. Suppose f.p.o.c. that there exists $\mathbf{x}$ such that $\nabla^2 f(\mathbf{x})$ is not PSD. Then, there exist $\lambda > 0$ and $\mathbf{p} \in \mathbb{R}^n$ such that $\langle \nabla^2 f(\mathbf{x})\mathbf{p}, \mathbf{p} \rangle = -\lambda$. As f is twice continuously differentiable, Taylor's theorem tells us that for all $\alpha > 0$ there exists $\gamma \in (0, 1)$ such that:

$$f(\mathbf{x} + \alpha\mathbf{p}) = f(\mathbf{x}) + \alpha \langle \nabla f(\mathbf{x}), \mathbf{p} \rangle + \frac{\alpha^2}{2} \langle \nabla^2 f(\mathbf{x} + \gamma\alpha\mathbf{p})\mathbf{p}, \mathbf{p} \rangle .$$

As $\langle \nabla^2 f(\mathbf{x})\mathbf{p}, \mathbf{p} \rangle = -\lambda$ and $\nabla^2 f$ is continuous, for all sufficiently small $\alpha > 0$:

$$\langle \nabla^2 f(\mathbf{x} + \gamma\alpha\mathbf{p})\mathbf{p}, \mathbf{p} \rangle \le -\frac{\lambda}{2} < 0.$$

But then also:

$$f(\mathbf{x} + \alpha\mathbf{p}) < f(\mathbf{x}) + \alpha \langle \nabla f(\mathbf{x}), \mathbf{p} \rangle ,$$

which is a contradiction, due to what we have proved in Q 6. □

Q 8. Let $f(\mathbf{x}) := \mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x}$, where $\mathbf{b}$ is the vector of all 1's and $\mathbf{A}$ is an $n \times n$ matrix defined by: $A_{ii} = 2$ for $1 \le i \le n$, $A_{i+1,i} = A_{i,i+1} = -1$, for $1 \le i \le n-1$ and $A_{n,1} = A_{1,n} = -1$. That is, $\mathbf{A}$ is defined as:

$$\mathbf{A} = \begin{bmatrix} 2 & -1 & 0 & 0 & \dots & 0 & -1 \\ -1 & 2 & -1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & 0 & -1 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 0 & 0 & 0 & \dots & -1 & 2 \end{bmatrix}.$$

Is f convex? Justify your answer. [15pts]

Solution. Observe that, $\forall \mathbf{x} \in \mathbb{R}^n$, $\nabla^2 f(\mathbf{x}) = \mathbf{A}$. We have discussed in class that $\nabla^2 f(\mathbf{x}) \succeq 0$ implies that f is convex (you can also prove this using Taylor's theorem). Thus, to show that f is convex, it suffices to prove that $\mathbf{A}$ is PSD. Take $\mathbf{x}$ to be any vector from $\mathbb{R}^n$. Then:

$$\langle \mathbf{A}\mathbf{x}, \mathbf{x} \rangle = \sum_{i=1}^{n-1} (x_i - x_{i+1})^2 + (x_1 - x_n)^2 \ge 0,$$

which proves that $\mathbf{A}$ is PSD. □

Q 9. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function that satisfies the following:

$$(\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n) : \quad f(\mathbf{y}) \ge f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{m}{2}\|\mathbf{y} - \mathbf{x}\|_2^2,$$

where $m > 0$ is some constant.

Prove that f cannot be Lipschitz continuous on the entire $\mathbb{R}^n$. Would it be possible for f to be Lipschitz continuous if we take the domain to be the unit Euclidean ball (i.e., the set: $\{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \le 1\}$)? Explain. [15pts]

Solution. It is not possible for f to simultaneously be m-strongly convex and L-Lipschitz continuous on the entire $\mathbb{R}^n$ for any $m, L > 0$ (m and L are finite). To see this, let $\mathbf{x}$ be such that $\nabla f(\mathbf{x}) \ne \mathbf{0}$ (such a point must exists, as the minimum of a strongly convex function is unique and the function is defined on the entire $\mathbb{R}^n$) and let $\mathbf{y} = \mathbf{x} + \alpha\nabla f(\mathbf{x})$, for some $\alpha > 0$. Strong convexity then gives us:

$$f(\mathbf{y}) - f(\mathbf{x}) \ge \alpha\Big(1 + \alpha\frac{m}{2}\Big)\|\nabla f(\mathbf{x})\|_2^2, \tag{2}$$

while Lipschitz continuity implies:

$$f(\mathbf{y}) - f(\mathbf{x}) \leq L\alpha \|\nabla f(\mathbf{x})\|_2. \tag{3}$$

But, for, e.g., $\alpha \geq \frac{2}{m} \frac{L}{\|\nabla f(\mathbf{x})\|_2}$, the lower bound on $f(\mathbf{y}) - f(\mathbf{x})$ from Eq. (2) is larger than the upper bound on $f(\mathbf{y}) - f(\mathbf{x})$ from Eq. (3), which is a contradiction.

If the feasible set were a unit ball $\mathcal{B}_2 = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_2 \leq 1\}$, then both conditions could be satisfied for

$$L \geq \max_{\mathbf{x},\mathbf{y} \in \mathcal{B}_2} \Big\{ f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{m}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \Big\}.$$

Note that the right-hand side of the last inequality is finite as f is real valued. It can be bounded above by: $\max_{\mathbf{x} \in \mathcal{B}_2} f(\mathbf{x}) + \max_{\mathbf{x} \in \mathcal{B}_2} \|\nabla f(\mathbf{x})\|_2 + \frac{m}{2}$. $\square$